Investigating mitochondrial DNA relationships in Neolithic Western Europe through serial coalescent simulations

Rivollat, M., Rottier, S., Couture, C., Pemonge, M.H., Mendisco, F., Thomas, M.G., Deguilloux, M.F. and Gerbault, P.

This is an author's accepted manuscript of an article published in the European Journal of Human Genetics, 28 December 2016. The final definitive version is available online at:

https://dx.doi.org/10.1038/ejhg.2016.180

SHORT REPORT

Title:
Investigating mitochondrial DNA relationships in Neolithic Western Europe through serial coalescent simulations

Running title:
Genetic relationships in Neolithic western Europe

Authors:
Maïté Rivollat[1], Stéphane Rottier[1], Christine Couture[1], Marie-Hélène Pemonge[1], Fanny Mendisco[1], Mark G. Thomas[2], Marie-France Deguilloux[1], Pascale Gerbault[2,3,4]

Affiliations:
[1] De la Préhistoire à l'Actuel, Culture, Environnement, Anthropologie – UMR 5199, CNRS, Université de Bordeaux, Bordeaux, France
[2] UCL Research Department of Genetics, Evolution and Environment, Darwin building, Gower Street, London WC1E 6BT, UK
[3] UCL Department of Anthropology, 14 Taviton Street, London WC1H 0BW, UK
[4] Department of Life Sciences, University of Westminster, 115 New Cavendish Street, London W1W 6UW, UK

Corresponding authors:
Maïté Rivollat (email : maite.rivollat@u-bordeaux.fr / maite87@live.fr; Tel: +33 (0)5 40 00 25 48; Fax: +33 (0)5 40 00 25 45); Pascale Gerbault (email: p.gerbault@ucl.ac.uk; Tel: +44 (0) 207 679 4397; Fax: +44 (0) 207 679 7193)

Conflict of interest:
The authors have nothing to disclose, no conflict of interest.

Abstract:

Recent ancient DNA studies on European Neolithic human populations have provided persuasive evidence of a major migration of farmers originating from the Aegean, accompanied by sporadic hunter-gatherer admixture into early Neolithic populations, but increasing towards the Late Neolithic. In this context, ancient mitochondrial DNA (mtDNA) data collected from the Neolithic necropolis of Gurgy (Paris Basin, France), the largest mtDNA sample obtained from a single archaeological site for the Early/Middle Neolithic period, indicate little differentiation from farmers associated to both the Danubian and Mediterranean Neolithic migration routes, as well as from western European hunter-gatherers. To test whether this pattern of differentiation could arise in a single unstructured population by genetic drift alone, we used serial coalescent simulations. We explore female effective population size parameter combinations at the time of the colonization of Europe 45 000 years ago and the most recent of the Neolithic samples analyzed in this study 5 900 years ago, and identify conditions under which population panmixia between hunter-gatherers/Early-Middle Neolithic farmers and Gurgy cannot be rejected. In relation to other studies on the current debate of the origins of Europeans, these results suggest increasing hunter-gatherer admixture into farmers' group migrating farther west in Europe.

Key words: genetic drift, European Neolithic, serial coalescent, ancient DNA, mtDNA

Introduction:

The introduction of farming into Europe around 8 600 years ago led to fundamental changes in subsistence strategy and social organization, and left signatures of

population turnover[1-4]. It is widely believed that farming spread into Europe from the Aegean along both Mediterranean and Danubian routes[3,5]. Recent archaeological[6] and palaeogenetic evidence[1,2,4,5,7] indicate a crucial role for migration, with only sporadic hunter-gatherer (HG) admixture into early Neolithic populations, but increasing towards the Late Neolithic[1,7,8]. However, these local inferences still permit spatiotemporal heterogeneity in HG admixture during the Neolithic in continental Europe.

In this context, the mtDNA diversity of the Gurgy "Les Noisats" site, located south of the Paris Basin and dated from 7 000 to 6 000 years ago, is striking since descriptive analyses[9] indicated affinities not only with early farmers associated with both the Danubian and Mediterranean migration routes but also with European HG. Notably, a relatively lower differentiation between Gurgy and European HG (F_{ST}=0.08) was observed when compared to other published levels of differentiation between Early Neolithic farmers and HG (e.g. F_{ST}=0.0923 (ref. 10); F_{ST}=0.163 (ref. 2)). This suggests complex admixture pattern between HG and farmer groups to shape Gurgy mtDNA diversity.

Previous mtDNA studies[2,10,11] have used serial coalescent simulations to test for genetic continuity between HG, Neolithic farmers and extant DNA samples from the same geographic region, and regularly concluded in genetic discontinuity between groups. We used a similar approach to address if the observed level of mtDNA differentiation between European HG, Neolithic farmer and Gurgy groups could be obtained under a panmictic population model with various combinations of effective population sizes. Our approach differs in three major aspects from previous studies[2,10,11]: first, we grouped the ancient mtDNA sequences according to subsistence strategy (HG or Neolithic farmers) and Neolithic context (Mediterranean/South-, Danubian/Central- or

82 Gurgy- farmers). Some of the sample groups are consequently contemporaneous and
83 can represent various regions. Second, we did not include modern population sample
84 into the comparison. Third, we extended the effective population size ranges used
85 previously[2,11] towards the lower bound to explore further demographic scenarios.
86
87 MATERIAL AND METHODS
88 We compiled 282 available ancient mtDNA HVR-I sequences
89 (NC_012920.1:m.16024_16380; Table S1). Following Rivollat et al. 2015, ancient
90 mtDNA data were partitioned into 4 sample groups: (i) Gurgy Les Noisats necropolis
91 (hereafter referred to as "Gurgy", n=39 sequences), (ii) Neolithic farmers from south
92 Europe (group "South-F", n=56, partitioned into 4 chronological sub-groups), (iii)
93 Neolithic farmers from central Europe ("Central-F", n=147, 5 sub-groups), and (iv)
94 hunter-gatherers ("HG", n=40, 16 sub-groups). Chronological sub-groups were defined
95 according to both shared geographic location and median calibrated C14 dates (see
96 Figure 1 and Figure 2). As a test statistic that measures the level of population
97 differentiation, we calculated six pairwise F_{ST} between the four groups (Figure 3) with
98 ARLSUMSTAT version 3.5.1.2 (ref. 12).
99 Following previous studies[2,11] we performed serial coalescent simulations under a single
100 panmictic population model with two demographic events: an initial colonization of
101 Europe 45 000 years ago of female effective population size N_{UP}, followed by
102 exponential growth or decline to the Neolithic transition in Western Europe 5 900 years
103 cal. BP of female effective population size N_N. Prior to N_{UP} we assume an ancestral
104 female effective population size N_A of 5 000, derived from the commonly used long-
105 term effective human population size of 10 000 individuals outside Africa[13] and

106 assuming a 1:1 female to male ratio. We explored 50 values for N_{UP} ranging from 1 to 5
107 000 and 50 values for N_N ranging from 10 to 100 000 (Table S2). We generated 50 000
108 mitochondrial genealogies of ancient HG and farmer sequences using fastsimcoal
109 version 2.5.1 (ref. 14) under each of the 2 500 N_{UP} - N_N combinations (Table S2). We
110 used a fixed mutation rate of 5×10^{-6}/bp/generation (ref. 15), assuming a 25 years
111 generation time. These simulated genealogies were used to compute expected pairwise
112 F_{ST} values for the six sample comparisons (Figure 2). We recorded the proportion of
113 simulated F_{ST} values that were greater than those observed per F_{ST} and parameter
114 combination (Figure 3).
115 We also tested if the six observed pairwise F_{ST} values as well as eight within sample
116 group statistic values (number of segregating sites and of pairwise differences) could be
117 recovered from simulations under this simple model by performing an approximate
118 Bayesian computation (ABC) -related approach[16] (see details in SI). We used the
119 rejection algorithm of the 'abc' package[17] available in R to retain the parameter
120 combinations that generated simulated pairwise F_{ST} the closest to the 6 observed values.
121 Even though we provide some effective population size estimates, we caution against
122 over-interpretation since there is likely insufficient information in the data to make
123 precise estimates.
124
125
126 Results and Discussion
127 Analyses indicate that for the six pairwise population group comparisons, some N_{UP} -
128 N_N combinations can result in simulated differentiation greater than the one observed
129 (grey area on Figure 3). Notably, results show that we cannot reject the possibility that

European HG, South-F, Central-F and Gurgy were sampled from a single panmictic population. Whereas these results may appear to contrast with previous studies that have used serial coalescent simulations to address local mtDNA population continuity between diachronic HG and farmers samples[2,11], we highlight that our analyses do not address 'population continuity' as defined in these studies. The grouping of diachronic samples may artificially reduce the level of differentiation that would be observed in case of significant mtDNA population structure. This grouping none-the-less allows us to investigate the genetic relationships between set of lineage samples associated with specific archaeological Neolithic contexts.

We confirmed that our panmictic population model generated simulated between and within population group diversity values close to the observed using an ABC-rejection algorithm (see SI and Figure S1). The 95% credible intervals estimated from the retained simulations are [5 – 3500] N_{UP} females and [200 – 7750] N_N females. These estimates concur with the observation that the parameter space for which a panmictic population model may hold is rather narrow (Figure 3). Most N_N values tested and compatible with the level of mtDNA differentiation observed are relatively low (10 to 200 females for the South-F and Central-F comparison, Figure 3). Noteworthy, some N_{UP} - N_N combinations imply a population decline that clearly contrasts with previous studies based on modern DNA data which have inferred female effective population size growth in Europe during the Holocene[18]. However, we were not constrained to simulate population expansion, since we did not consider modern DNA data in our analyses. Moreover, a Holocene population decline in Europe corroborates recent Y chromosome data[18] and various archaeological evidence support demographic fluctuation of Neolithic populations[19,20].

Our results indicate that a simple panmictic population model can account for the mtDNA differentiation observed between European HG and Early/Middle Neolithic farmers; a larger proportion of the HG - Gurgy explored parameter space failed to reject panmixia. This result suggests increasing HG admixture into farmers' group migrating farther west in Europe. Similarly, we note that a larger proportion of the explored parameter space fails to reject panmixia when comparing Gurgy and South-F than when comparing Gurgy and Central-F. Thus, our results seem to support Gurgy as the most ancient Neolithic sample studied so far with appreciable admixture between pre-Neolithic HG and Early/Middle Neolithic farmers from both streams of Neolithization in Europe (with a suspected higher participation of Mediterranean farmers).

As with any model, the one we test here has a few assumptions that may not hold, e.g. N_A of female to male ratio of 1 (ref. 18) and no population structure in any of the four groups[5]. Moreover, the panmictic population model proposed would need to be compared against alternatives (e.g. ref. 11). Such a simple panmictic population model nevertheless lays the ground for building more complex ones[17]. Notably, a serial coalescent approach coupled with ABC would allow estimation of the possible contribution of each of the three population groups (HG, Mediterranean and Central Europe farmers) in shaping Gurgy mitochondrial diversity.

Acknowledgements:

This work was supported by a ministerial grant from the Research National Agency as a program of prospects investments (ANR-10-LABX-52, DHP project; dir: SR; Université Bordeaux 1, LaScArBx-ANR; 2012-14; and ANR-10-IDEX-03-02 to MR

for work in UCL) and a Leverhulme Programme grant to A.M. Migliano (UCL Anthropology) and M.G. Thomas (RP2011-R-045). The authors acknowledge the use of the UCL Legion High Performance Computing Facility (Legion@UCL), and associated support services, in the completion of this work.

Supplementary information is available at *European Journal of Human Genetics*' website.

References

1 Haak, W., Lazaridis, I., Patterson, N. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 2015; **552**, 207-211.

2 Bramanti, B., Thomas, M. G., Haak, W. *et al.* Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 2009; **326**, 137-140, doi:10.1126/science.1176869.

3 Childe, V. G. *The Dawn of European Civilization*. Vol. 674 Kegan Paul, 1925; 328.

4 Lazaridis, I., Patterson, N., Mittnik, A. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 2014; **513**, 409-413.

5 Hofmanová, Z., Kreutzer, S., Hellenthal, G. *et al.* Early farmers from across Europe directly descended fron Neolithic Aegeans. *Proceedings of the National Academy of Sciences* in press.

6 Bocquet-Appel, J.-P., Naji, S., Vander Linden, M. & Kozlowski, J. K. Detection of diffusion and contact zones of early farming in Europe from the space-time distribution of 14C dates. *Journal of Archaeological Science* 2009; **36**, 807-820.

7 Brandt, G., Haak, W., Adler, C. J. *et al.* Ancient DNA Reveals Key Stages in the Formation of Central European Mitochondrial Genetic Diversity. *Science* 2013; **342**, 257-261, doi:10.1126/science.1241844.

8 Bollongino, R., Nehlich, O., Richards, M. P. *et al.* 2000 years of parallel societies in Stone Age Central Europe. *Science* 2013; **342**, 479-481, doi:10.1126/science.1245049.

9 Rivollat, M., Mendisco, F., Pemonge, M.-H. *et al.* When the Waves of European Neolithization Met: First Paleogenetic Evidence from Early Farmers in the Southern Paris Basin. *PLoS One* 2015; **10**, e0125521, doi:10.1371/journal.pone.0125521.

10 Haak, W., Balanovsky, O., Sanchez, J. J. *et al.* Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS biology* 2010; **8**, e1000536, doi:10.1371/journal.pbio.1000536.

11 Rasteiro, R. & Chikhi, L. Female and male perspectives on the neolithic transition in Europe: clues from ancient and modern genetic data. *PLoS One* 2013; **8**, e60944, doi:10.1371/journal.pone.0060944.

12 Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* 2010; **10**, 564-567.

13 Consortium, T. G. P. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**, 1061-1073.

14 Excoffier, L. & Foll, M. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 2011; **27**, 1332-1334.

15 Soares, P., Ermini, L., Thomson, N. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *The American Journal of Human Genetics* 2009; **84**, 740-759.

16 Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* 2002; **162**, 2025-2035.

17 Csilléry, K., Blum, M. G., Gaggiotti, O. E. & François, O. Approximate Bayesian computation (ABC) in practice. *Trends in Ecology & Evolution* 2010; **25**, 410-418.

18 Karmin, M., Saag, L., Vicente, M. *et al.* A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research* 2015; **25**, 459-466.

19 Shennan, S., Downey, S. S., Timpson, A. *et al.* Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature communications* 2013; **4**, 2486.

20 Wirtz, K. W. & Lemmen, C. A global dynamic model for the Neolithic transition. *Climatic Change* 2003; **59**, 333-367.

247

248

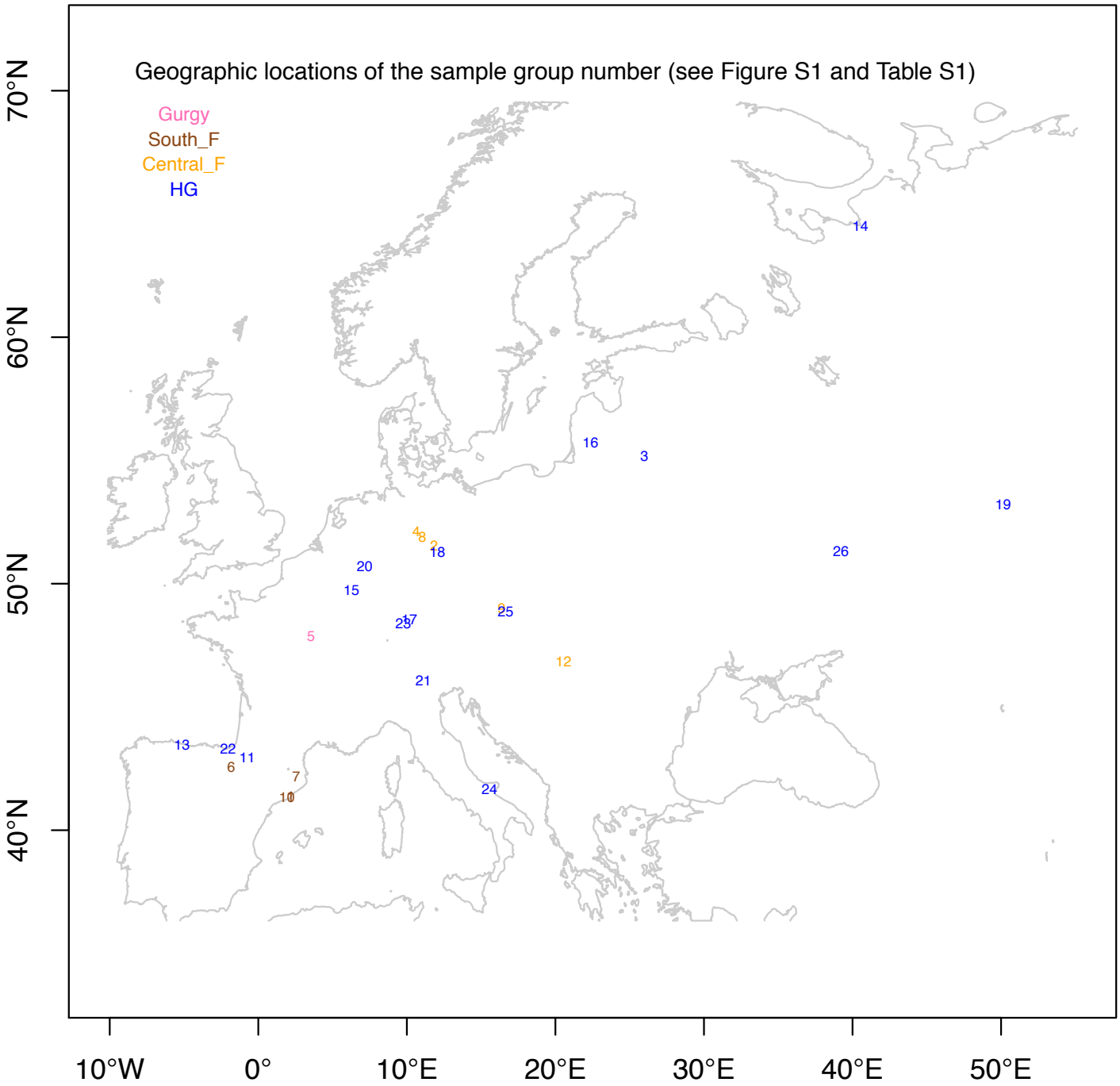
‘Titles and legends to figures’:

Figure 1: Geographic locations of the sites sampled for ancient mtDNA in the current analysis. 1 refers to the most recent sample group in the gene genealogy and 26 to the most ancient. Groups were categorized based on common median C14 dates, archaeological context and geographic location; one group can consequently be sampled from more than one geographic location; for those groups with more than one location, only the location of the largest sample size group is displayed on the map (details in Figure 2 and Table S1).

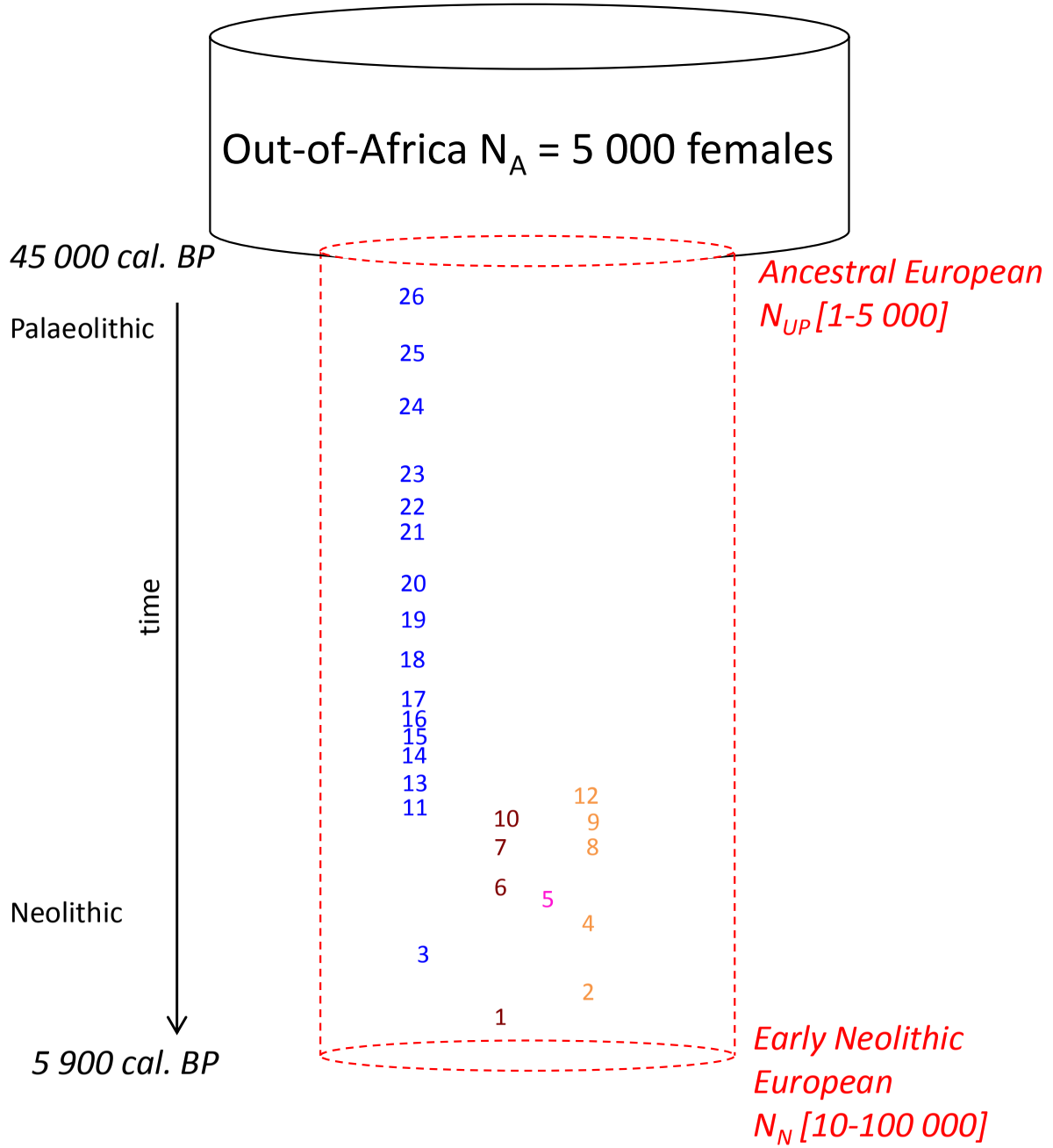
Figure 2: Demographic model simulated with the serial coalescent. Time is the median calibrated C14 years before present (cal BP) backward in time from ‘t0’ and expressed in generations. ‘t0’ refers to 5 900 years cal. BP, the median C14 date of the youngest ancient mtDNA sample. Groups are numbered backward in time from the most recent to the most ancient. The dashed red cylinder shows constant population size between N_{UP} and N_N, but the simulated population can undergo expansion or decline depending on the combinations of these parameter values.

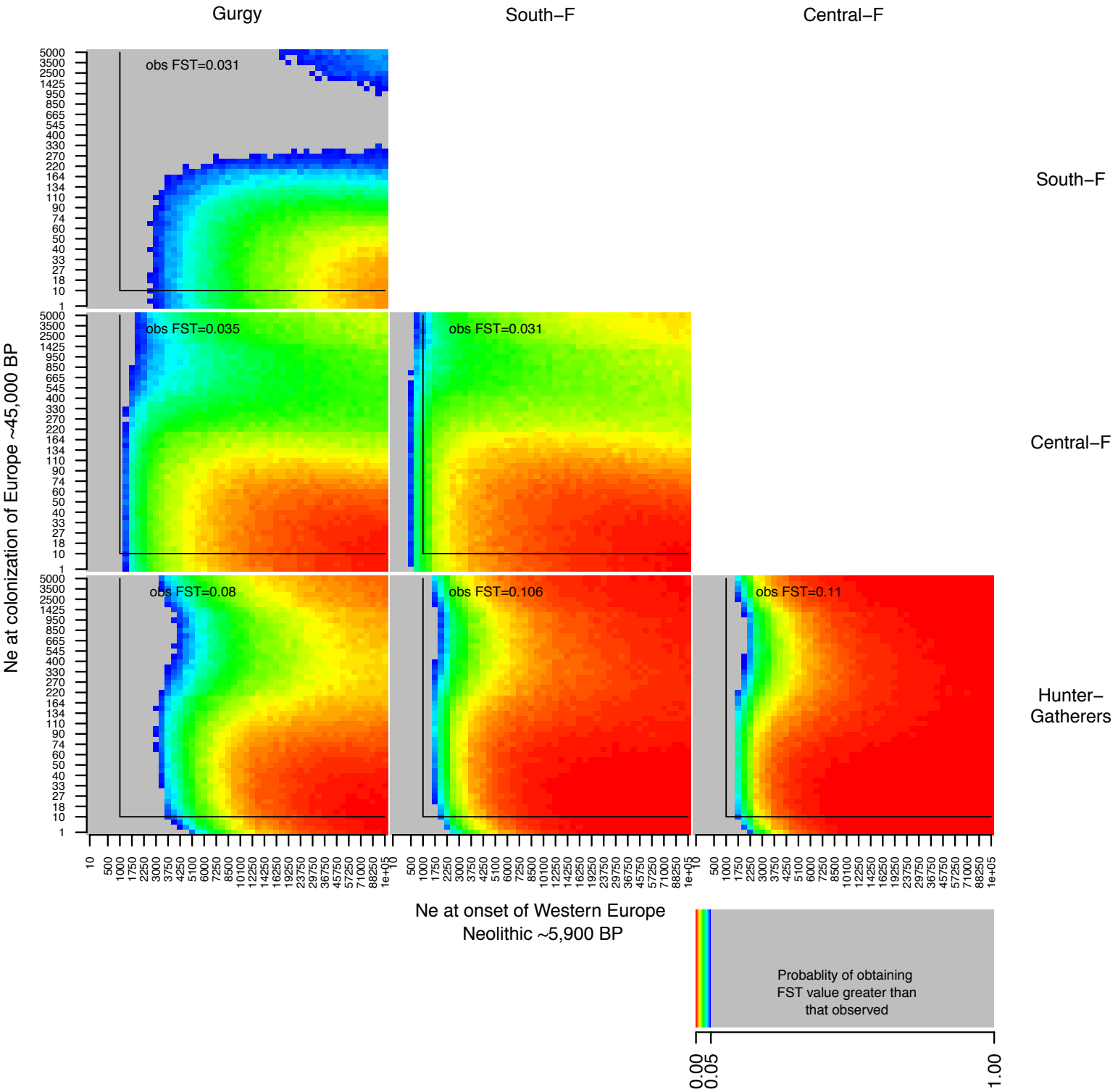
Figure 3: Probability of obtaining simulated F_{ST} value greater than that observed for the six pairwise population groups compared (see text for details). Corresponding observed pairwise F_{ST} are shown in the top left corner of each grid. The 50x50 grids show values of assumed effective population size N_N on the x-axis and values of parameterized N_{UP} on the y-axis (note that 25 values are shown on each axis for clarity, see Table S2). The top right area delimited by vertical and horizontal black lines outline N_N and N_{UP} ranges, respectively, used in comparable studies[2,11]. Grey shows proportions of observed F_{ST} greater than observed (proportion > 0.05), for which panmixia cannot be rejected. Color-scale represents

significance level from blue (proportion lower or equal to 0.05) to red (proportion close to 0). Proportions were obtained over 50 000 simulated pairwise F_{ST} per combination of N_N and N_{UP} value.


Geographic locations of the sample group number (see Figure S1 and Table S1)
Gurgy
South_F
Central_F
HG
70°N
60°N
50°N
40°N
10°W
0°
10°E
20°E
30°E
40°E
50°E
14
16
3
19
4
8
2
18
26
20
15
25
17
25
5
12
21
13
22
11
6
7
10
24

Population name (total n)	group NB	Sample Size (n)	Time (gen back)	Median date (cal. BP)
Gurgy	5	39	24	6500
South_F (56)	1	3	0	5900
	6	36	32	6700
	7	7	42	6950
	10	10	54	7250
Central_F (147)	2	32	4	6000
	4	17	20	6400
	8	88	44	7000
	9	7	52	7190
	12	3	71	7685
HG (40)	3	3	14	6250
	11	1	64	7500
	13	2	76	7800
	14	11	86	8050
	15	1	96	8300
	16	2	96	8300
	17	2	110	8650
	18	1	116	8800
	19	2	144	9500
	20	2	312	13700
	21	1	324	14000
	22	3	344	14500
	23	1	378	15350
	24	4	764	25000
	25	3	1010	31155
	26	1	1343	39475




Gurgy
South−F
Central−F
South−F
Central−F
Hunter−
Gatherers
obs FST=0.031
obs FST=0.035
obs FST=0.031
obs FST=0.08
obs FST=0.106
obs FST=0.11
Ne at colonization of Europe ~45,000 BP
Ne at onset of Western Europe
Neolithic ~5,900 BP
Probablity of obtaining
FST value greater than
that observed
0.00
0.05
1.00